Titanic Data Visualization

In [3]:
```
import pandas as pd
import numpy as np
import math
import matplotlib.pyplot as mplot
from matplotlib import cm
```

In [7]:
```
#Read data
df = pd.read_table("E:\\Data Projects\\titanic\\titanic.txt", sep=",")
```

In [8]:
```
display(df)
```

	Survived	Passenger Class	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare in British Pounds
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000
4	0	3	Mr. William Henry Allen	male	35.0	0	0	8.0500
...	...	...	...	...	...	...	...	...
882	0	2	Rev. Juozas Montvila	male	27.0	0	0	13.0000
883	1	1	Miss. Margaret Edith Graham	female	19.0	0	0	30.0000
884	0	3	Miss. Catherine Helen Johnston	female	7.0	1	2	23.4500
885	1	1	Mr. Karl Howell Behr	male	26.0	0	0	30.0000
886	0	3	Mr. Patrick Dooley	male	32.0	0	0	7.7500

887 rows × 8 columns

In [9]:
```
df = df.rename(columns={'Fare in British Pounds' : 'fare', 'Siblings/Spouses Aboard' : 'SibSpouse',
                        'Parents/Children Aboard' : "ParChild",
                       'Passenger Class' : 'class'})
#converting column names for ease of use
display(df)
```

	Survived	class	Name	Sex	Age	SibSpouse	ParChild	fare
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000
4	0	3	Mr. William Henry Allen	male	35.0	0	0	8.0500
...	...	...	...	...	...	...	...	...
882	0	2	Rev. Juozas Montvila	male	27.0	0	0	13.0000
883	1	1	Miss. Margaret Edith Graham	female	19.0	0	0	30.0000

	Survived	class	Name	Sex	Age	SibSpouse	ParChild	fare
884	0	3	Miss. Catherine Helen Johnston	female	7.0	1	2	23.4500
885	1	1	Mr. Karl Howell Behr	male	26.0	0	0	30.0000
886	0	3	Mr. Patrick Dooley	male	32.0	0	0	7.7500

887 rows × 8 columns

In [10]:
```
#finding average price of first class fare
fc = df.loc[(df['class']==1)]
fcfaremean = fc['fare'].mean()
fcfaremean = fcfaremean * 1.28
print("Average price of first class fare, in American $:", fcfaremean)
```

```
Average price of first class fare, in American $: 107.7179999999999
```

In [11]:
```
#passengers over 20 with siblings
ov20wsib = df.loc[(df['Age']>20) & (df['SibSpouse']>1)]
#display(ov20wsib)
ov20index = ov20wsib.index
print("Passengers over 20 with siblings:",len(ov20index))
```

```
Passengers over 20 with siblings: 23
```

In [12]:
```
#average age of those who did not live
didntlive = df.loc[(df['Survived']==0)]
dlmedian = didntlive['Age'].median()
print('Median age of those who didn\'t live:', dlmedian, 'years old.')
```

```
Median age of those who didn't live: 28.0 years old.
```

In [13]:
```
#percent of male passengers who survived
totpassindex = df.index
totpass = len(totpassindex)
maleslived = df.loc[(df['Sex']=='male') & (df['Survived']==1)]
mlindex = maleslived.index
mllength = int(len(mlindex))

print("Percentage of males who lived:", ((mllength / totpass))*100,"%")
```

```
Percentage of males who lived: 12.288613303269448 %
```

In [14]:
```
#percent of female passengers who survived
totpass = len(df)
fmlived = df.loc[(df['Sex']=='female') & (df['Survived']==1)]
flindex = fmlived.index
fllength = int(len(flindex))

print("Percentage of females who lived:", ((fllength / totpass))*100,"%")
```

```
Percentage of females who lived: 26.268320180383313 %
```

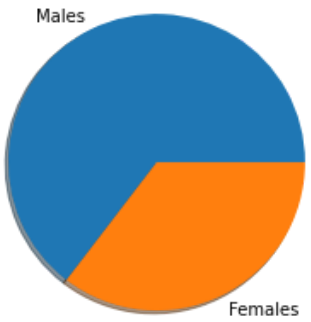
In [15]:
```
#percent of ppl in 3rd class who survived vs percent of ppl in 1st class who survived
tcsurv = df.loc[(df['class']==3) & (df['Survived']==1)]
tcsurvlen = len(tcsurv)
fcsurv = fc.loc[(df['class']==1) & (df['Survived']==1)]
fcsurvlen = len(fcsurv)

print("The percentage of people in 3rd class who lived:",((tcsurvlen/totpass)*100),"%")
print("The percentage of people in 1st class who lived:",((fcsurvlen/totpass)*100),"%")
```

```
The percentage of people in 3rd class who lived: 13.416009019165728 %
The percentage of people in 1st class who lived: 15.332581736189402 %
```

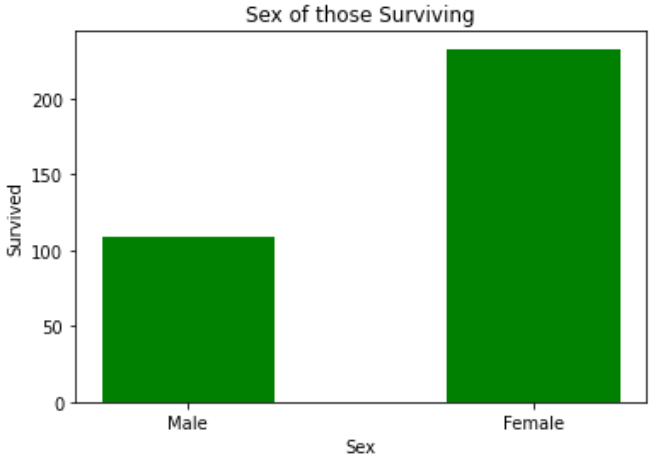
In [16]:
```
males = df.loc[(df['Sex']=='male')]
mlength = len(males)
females = df.loc[df['Sex']=='female']
flength = len(females)

scount = [['males', mllength], ['females', fllength]]
```

In [17]:
```
#sframe = pd.DataFrame(scount, columns = ['Sex', 'Count'])
#sframe.groupby(['Sex']).sum().plot(kind='pie', y='Count')
```

In [18]:
```
pielabs = 'Males', 'Females'
sizes = [mlength, flength]

spie = mplot.pie(sizes, labels=pielabs, shadow=True)
```

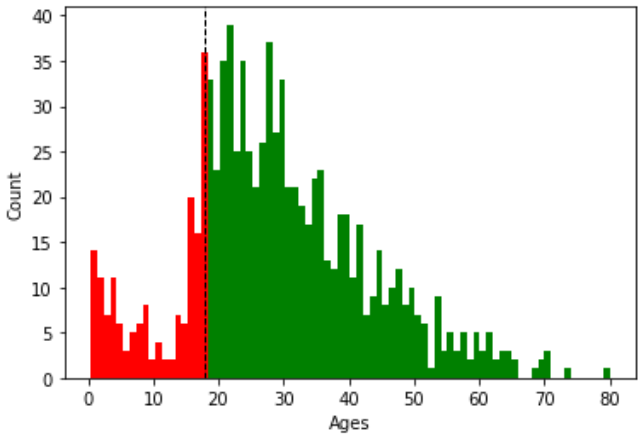


In [19]:
```
survivalists = {'Male' : mllength, 'Female': fllength}
sx = ['Male', 'Female']
scounts = [mllength, fllength]

xes = [i for i, _ in enumerate(sx)]

mplot.bar(xes, scounts, color='green', width =0.5)
mplot.xlabel("Sex")
mplot.ylabel("Survived")
mplot.title("Sex of those Surviving")
mplot.xticks(xes,sx)
mplot.show()

#survframe = pd.DataFrame(survivalists, columns=['Sex', 'Count'])
#survframe.groupby(['Sex']).sum().plot(kind='bar', y='Count', title='Those Who Survived')
```



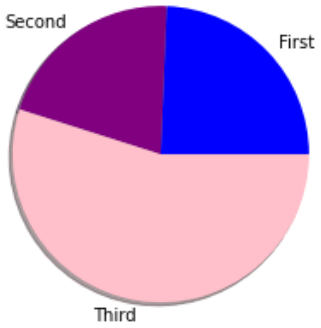
In [20]:
```
_, _, histplot = mplot.hist(df['Age'], bins = 80, color='green')
         # , bins = histbins)
```

```
    #To show the total passengers under the age of 18
for bar in histplot:
    if bar.get_x() < 18:
        bar.set_facecolor("red")

mplot.xlabel("Ages")
mplot.ylabel("Count")
mplot.axvline(x=18, linestyle='--', linewidth=1, color='black')

mplot.show()
```



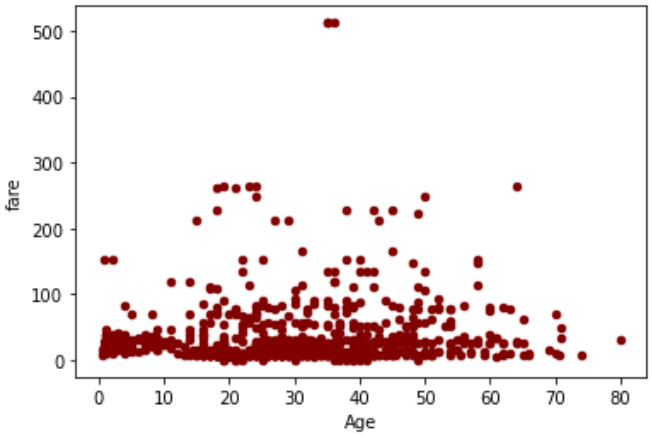
In [21]:

```
#independent graph 1:  pie chart of class
fc = df.loc[(df['class']==1)]
fc = int(len(fc))
sc = df.loc[(df['class']==2)]
sc = int(len(sc))
tc = df.loc[(df['class']==3)]
tc = int(len(tc))

cl_lab = ['First', 'Second', 'Third']
cl_sizes = [fc, sc, tc]

cs= ['blue','purple','pink']

cpie = mplot.pie(cl_sizes, labels=cl_lab, shadow=True, colors=cs)
```



In [22]:

```
#scatterplot of age vs. fare price
ax1 = df.plot.scatter(x='Age', y='fare', c='maroon')
```



In [23]:

```
df['fare'] = pd.to_numeric(df['fare'])

_, _, farehist = mplot.hist(df['fare'], bins = 12, color='purple')

mplot.show()
```